






# Comparison of denoising tools for the reconstruction of nonlinear multimodal images

ROLA HOUHOU,<sup>1,2</sup>  ELSIE QUANSAH,<sup>1,2</sup> TOBIAS MEYER-ZEDLER,<sup>1,2</sup> MICHAEL SCHMITT,<sup>1</sup>  FRANZISKA HOFFMANN,<sup>3</sup>  ORLANDO GUNTINAS-LICHIUS,<sup>3</sup> JÜRGEN POPP,<sup>1,2</sup>  AND THOMAS BOCKLITZ<sup>1,2,4,\*</sup> 

<sup>1</sup>*Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich Schiller University, Helmholtzweg 4, 07743 Jena, Germany*

<sup>2</sup>*Leibniz Institute of Photonic Technology (Member of Leibniz Health Technologies), Albert-Einstein-Straße 9, 07745 Jena, Germany*

<sup>3</sup>*Department of Otorhinolaryngology, Institute of Phoniatry/Pedaudiology, Jena University Hospital, Jena, Germany*

<sup>4</sup>*Institute of Computer Science, Faculty of Mathematics, Physics and Computer Science, University Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany*

\**Thomas.bocklitz@uni-jena.de*

**Abstract:** Biophotonic multimodal imaging techniques provide deep insights into biological samples such as cells or tissues. However, the measurement time increases dramatically when high-resolution multimodal images (MM) are required. To address this challenge, mathematical methods can be used to shorten the acquisition time for such high-quality images. In this research, we compared standard methods, e.g., the median filter method and the phase retrieval method via the Gerchberg-Saxton algorithm with artificial intelligence (AI) based methods using MM images of head and neck tissues. The AI methods include two approaches: the first one is a transfer learning-based technique that uses the pre-trained network DnCNN. The second approach is the training of networks using augmented head and neck MM images. In this manner, we compared the Noise2Noise network, the MIRNet network, and our deep learning network namely incSRCNN, which is derived from the super-resolution convolutional neural network and inspired by the inception network. These methods reconstruct improved images using measured low-quality (LQ) images, which were measured in approximately 2 seconds. The evaluation was performed on artificial LQ images generated by degrading high-quality (HQ) images measured in 8 seconds using Poisson noise. The results showed the potential of using deep learning on these multimodal images to improve the data quality and reduce the acquisition time. Our proposed network has the advantage of having a simple architecture compared with similar-performing but highly parametrized networks DnCNN, MIRNet, and Noise2Noise.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

## 1. Introduction

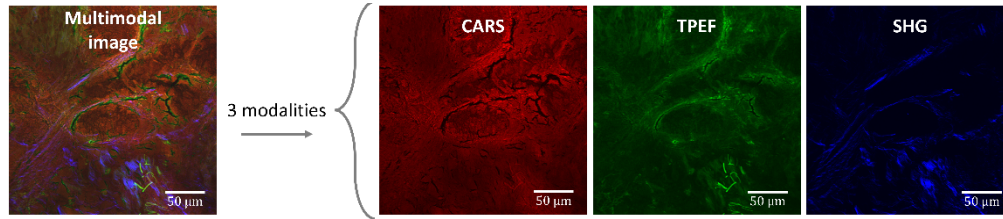
Medical imaging is an important and active area of research with the potential to significantly improve disease diagnosis and patient treatment. For decades, medical imaging modalities, e.g., X-ray [1], ultrasound imaging [2–4], and computerized tomography (CT) [1,5], have served as important tools to assist physicians in making their diagnostic decisions. Although several new especially optical imaging technologies have been developed in the last decades, their adoption in healthcare systems is still minimal. Nonlinear optical techniques, e.g., coherent anti-Stokes Raman scattering (CARS) [6], two-photon excited fluorescence (TPEF) [7], and second-harmonic generation (SHG) [8], and linear optics techniques, e.g., fluorescence lifetime imaging (FLIM) [9] are capable of measuring detailed information about the chemical composition and morphology of tissue sections with high spatial resolution and in a non-altering manner. In particular,

the simultaneous combination of two or more of these optical spectroscopic methods, called multimodal imaging (MM), allows for maximizing the obtained chemical and morphological information of the measured tissues [10–15]. For instance, Vogler et al. [8] presented a microscopic experiment that combines three nonlinear optical techniques; CARS, TPEF, and SHG, and shows how different kinds of molecules and different contrast mechanisms can be obtained in one image measurement. In detail, CARS measurements explore the molecular distribution like proteins and lipids, SHG measurements highlight collagen fiber distribution in the sample and TPEF measurements identify specific molecules like keratin and NADP(H). The combination of these three modalities is considered a label-free and non-destructive approach that is very useful for in vivo studies [16]. The multimodal imaging approach provides high-quality (HQ) images, but the acquisition of such high-quality images requires a relatively long acquisition process in comparison with low-quality images because photon shot noise is the prominent noise source in nonlinear imaging techniques. Mechanical methods such as using a faster motor [17] or time-stretching techniques [18] have shown great promise in improving the speed and performance of multimodal imaging systems, however, these methods have certain limitations. For instance, although using a faster motor can reduce scan times it may generate more heat which can potentially degrade the quality of the image. On the other hand, time-stretching techniques can increase the time resolution of imaging systems, but they may also introduce noise and distortions to the image. Additionally, the imaging system in this study uses laser scanning and only shifts the sample when jumping from tile to tile, so the measurement time is limited by the detector, not the scanning speed. Hence, the faster MM imaging required for real-time monitoring leads to an increase in the noise level of the images, which degrades their quality and affects the identification of tissues or their associated diseases, or abnormalities.

In addition to experimentally acquiring HQ images by increasing the acquisition time, image denoising is a fundamental preprocessing technique that can remove noise from images but may result in the loss of relevant information [19–21]. Consequently, the trade-off between fast imaging and a suitable denoising method needs to be balanced and optimized for an effective diagnostic imaging tool. The denoising algorithms vary from basic digital image filters to iterative reconstruction techniques. Therefore, choosing a suitable denoising method is not simple, and the restored images should maintain the following properties [20]. First, the details and edges that are critical to detect malignant tissue should be preserved. This means that the denoising algorithms should not produce artifacts and the recovered images should be similar to the original image. In addition, the algorithm should be computationally efficient and have low complexity, which is a prerequisite in medical applications that require immediate results. Finally, the denoising algorithms should not depend on vast amounts of data, which is not practical or readily accessible in medical imaging.

Apart from the standard image denoising methods, deep learning featured a high potential for denoising and showed outstanding performance, especially in the processing of natural images and various medical imaging techniques, e.g., ultrasound imaging [2–4], CT scan [1,5], fluorescence microscopy [22], and CARS endoscopy [6]. Therefore, we evaluated deep learning methods on the multimodal images that comprise CARS, TPEF, and SHG modalities and compared them with the following standard techniques; the median filter (MF) method and the phase retrieval method via Gerchberg-Saxton (GS) [23–27]. An example of a MM image is visualized in Fig. 1, where the CARS, TPEF, and SHG modalities are represented as the red, green, and blue channels, respectively. In this manuscript, we used two deep learning approaches. The first approach is a transfer learning-based method [28] in which we used the pre-trained network, namely DnCNN [29] directly to reconstruct the improved images. The second approach is to train a network using augmented neck and tissue MM images. In this context, we used two well-known architectures; the Noise2Noise (N2N) [30] and the MIRNet [31,32] architectures, in addition to our deep learning network that we referred to as incSRCNN. The incSRCNN network consists of a simple

architecture derived from the super-resolution convolution neural network (SRCNN) [33,34] with a small trick in the first layer that was inspired by the inception network [35]. In this manuscript, we briefly explain all the methods at the beginning and then describe the data and workflow. We then discuss the reconstruction of synthetic and experimental low-quality images using the GS algorithm, the MF method, the DnCNN, N2N, MIRNet, and incSRCNN networks. Afterward, a generalizability section is presented with two different analyses. Finally, we summarize our results in the conclusion section.



**Fig. 1.** An example of a multimodal image consisting of the three modalities CARS of the CH<sub>2</sub> stretching vibration at 2850 cm<sup>-1</sup>, TPEF, and SHG, as the red, green, and blue channels, respectively, is given.

## 2. Method

### 2.1. Direct methods: median filter, GS algorithm, and pre-trained DnCNN network

This section briefly explains the implemented methods, grouped into a description of the classical methods and deep learning methods. First, the median filter with a  $3 \times 3$  kernel size is used by computing the median value of the input image under the kernel window [36]. Then, the phase retrieval problem is implemented since it is applied to many phase-based denoising problems [37–40]. Several well-known phase retrieval algorithms exist, e.g., hybrid input-output (HIO) and Gerchberg-Saxton (GS). We focused on applying GS [23,25] to the MM images since most of the other error reduction-based techniques represent a derived version of the GS algorithm. Briefly, GS is the recovery of the phase using the measured image and the source object. It is considered an error-reduction algorithm that iteratively calculates the error until it converges. The GS algorithm is shown in Fig. 2, and it is applied independently on each channel where the phase and the modified amplitudes are determined iteratively, enabling image reconstruction. Its input represents both the amplitudes of the sampled image  $\sqrt{x}$  and a Gaussian estimation of the diffraction plane intensity  $X$ . First, an initial phase  $\varphi_0$  in the object plane is used by generating randomly uniform numbers between  $-\pi$  and  $\pi$ . At iteration  $k$ , the initial field in the object plane is calculated using Eq. (1).

$$z_k = \sqrt{x} \exp(i\varphi_{k-1}) \quad (1)$$

The phase distribution in the target plane  $\phi_k$  is then calculated via the fast Fourier transform (FFT), as shown in Eq. (2).

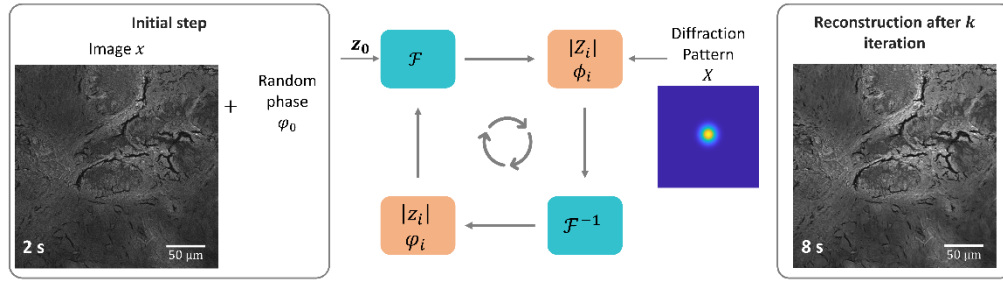
$$\phi_k = \arg(FFT(z_k)) \quad (2)$$

Equation (3) combines the phase distribution in the target plane with the target intensity  $\sqrt{X}$  and finally, the phase in the object plane  $\varphi_k$  is recovered by using Eq. (4).

$$A_k = \sqrt{X} \exp(i\phi_k) \quad (3)$$

$$\varphi_k = \arg(FFT(A_k)). \quad (4)$$

Apart from the classical methods used in image denoising, artificial intelligence (AI) based methods have widely been used for restoring images, especially in computer vision and medical



**Fig. 2.** The workflow of the GS algorithm. First, the LQ image with a random phase was fed to the algorithm, and after  $k$  iteration, the high-quality image was constructed. The GS algorithm depends on an estimation of the source object, which is unknown, and therefore Gaussian estimation was used.

imaging, e.g., X-ray, CT imaging, and ultrasound scans. In artificial intelligence, high-quality images, which represent improved images in terms of signal-to-noise ratio (SNR), can be acquired by transfer learning or directly constructing deep learning techniques. Transfer learning [28] consists of using knowledge obtained from one task and transferring it to another related task. The direct deep learning method, on the other hand, trains a neural network with specific architecture using the available data, optimizing the parameter during training. In this manuscript, we evaluated both approaches on the MM images.

First, the pre-trained neural network, the denoising convolutional neural network (DnCNN), was used as a transfer learning tool. DnCNN was trained on natural images to correct noise and artifacts in corrupted images [29]. Briefly, DnCNN is a pre-trained network that outputs the residual image, i.e., the difference between the noisy observation and the latent clean image, instead of predicting the denoised image. The architecture of this network is an adapted version of the VGG network [41] that is suitable for the image denoising task. Formally, the averaged mean squared error calculated in Eq. (5) between the desired residual images and estimated ones from noisy input

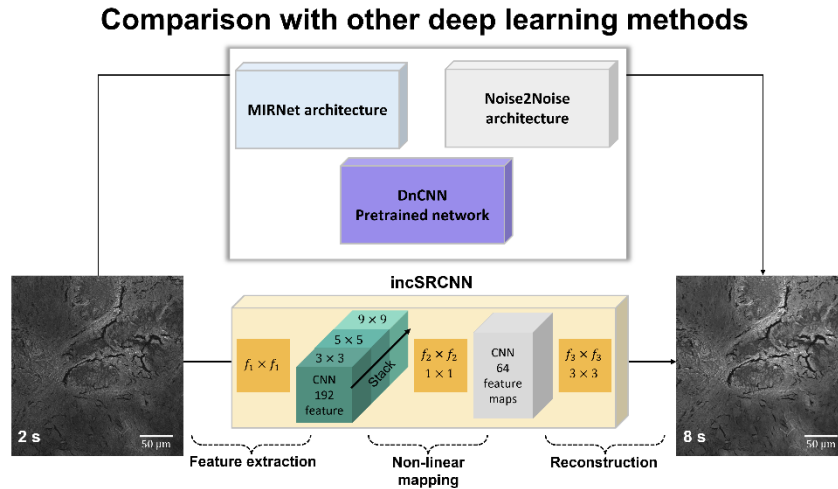
$$l(\theta) = \frac{1}{2N} \sum_{i=1}^N \|\mathfrak{R}(y_i; \theta) - (y_i - x_i)\|_F^2, \quad (5)$$

can be adopted as the loss function to learn the trainable parameters in DnCNN.  $\mathfrak{R}(y)$  represents the residual mapping and  $\{(y_i, x_i)\}_{i=1}^N$  represents  $N$  noisy-clean training image patch pairs. In a nutshell, the DnCNN model has two main features: the residual learning formulation is adopted to learn  $\mathfrak{R}(y)$ , and batch normalization is incorporated to speed up training and boost the denoising performance.

## 2.2. Trained networks: incSRCNN, N2N, and MIRNet

Then, we constructed and trained a simple network which is a modified version of the super-resolution convolutional neural network (SRCNN) [33,34], namely incSRCNN. Our architecture is inspired by both the inception and the SRCNN networks, therefore we call it incSRCNN. The architecture of this network is shown in Fig. 3. Like the SRCNN, the proposed network consists of three layers; however, it is implemented as a denoising task that outputs the same input size. The input image is convolved in the first layer with three different kernel sizes 3, 5, and 9 into 192 feature maps. The second layer then applies a  $1 \times 1$  kernel to condense to 64 feature maps. Finally, the third layer uses a  $3 \times 3$  kernel to construct the output image. All layers involve the ReLu activation function. We used the mean absolute error as a loss function between the original HQ image and the output from the trained networks and the weights in the network layers are updated using the Adam optimizer with a learning rate equal to  $3e^{-4}$ .





**Fig. 3.** The transfer learning-based approach via DnCNN and the trained deep learning networks via Noise2Noise, MIRNet, and our proposed deep learning networks (incSRCNN). On top of the figure, we used a pre-trained network, DnCNN, to predict MM images with higher quality. Moreover, the Noise2Noise and The MIRNet networks are trained using augmented neck and head tissue images. The architecture of our proposed network, incSRCNN, is shown at the bottom. This network represents a modified version of the SRCNN and is inspired by the inception network. Initially, the first layer convolves the input image with different kernel sizes into 192 feature maps. The second layer then applies a  $1 \times 1$  kernel to condense to 64 feature maps. Finally, the third layer uses a  $3 \times 3$  kernel to construct the output image.

Afterward, we aimed to compare our simple architecture with more complex ones. Therefore, we chose well-known networks: the Noise2Noise (N2N) and the MIRNet architecture which are usually implemented for denoising tasks. Briefly, N2N and MIRNet architectures consist of deep convolutional neural networks (CNN) layers. The N2N network learns to remove noise from a noisy image by training on pairs of noisy images, effectively learning to denoise without ever seeing a clean image. On the other hand, the MIRNet network uses a multi-scale architecture to capture both local and global image features, and incorporates a feature fusion module to combine information from different scales (we refer readers for more details about N2N and MIRNet to the Ref. [30] and the Refs. [31,32], respectively).

### 3. Data acquisition, description, and workflow

#### 3.1. Data acquisition and description

The data used for developing the denoising method has been acquired using a laser scanning microscope (LSM510, Zeiss, Germany) equipped with a ps-laser system for coherent anti-Stokes Raman scattering (CARS), second harmonic generation (SHG), and two-photon excited fluorescence (TPEF) microscopy as described in detail previously [42]. Briefly, the sample is illuminated with two spatially and temporally synchronized laser pulse trains of ps-pulse duration. The difference frequency of both lasers matches the symmetric CH<sub>2</sub> stretching vibration at  $2850 \text{ cm}^{-1}$ . The pump laser is operating at 672.5 nm, the Stokes laser is at 832 nm. The specimen is illuminated through a 20x planapochromatic objective (Zeiss, Germany, NA = 0.8) using a 50 mW pump and 70 mW of Stokes power. CARS and SHG signals are collected and detected in forward direction by PMT detectors. The signals are split by a 514 nm dichroic longpass mirror. The CARS signal is detected using a 550 nm bandpass filter, the SHG signal using a 415 nm

bandpass filter. The TPEF signal is collected in epi-direction through the illumination objective and reflected by a 600 nm longpass dichroic mirror to the PMT detector. In front of the PMT the TPEF signal is filtered using a 650 nm shortpass filter and a 458/64 nm bandpass filter (both Semrock, USA). All analyzed images have been acquired using 1.6  $\mu$ s pixel dwell time, a field of view of 450  $\mu$ m and 512 pixels length. For HQ images 16 frames have been averaged, for LQ images four frames averaging was applied.

The data represent the head and neck tissue of a mouse, with ten positions measured using the nonlinear multimodal imaging technique. The nonlinear multimodal imaging combines three modalities that are simultaneously excited using a 672.5 nm pump and 832 nm Stokes and detected at 550 nm (CARS), 458 nm (TPEF), and 415 nm (SHG). In this manuscript, we utilized high-quality (HQ) and (experimental) low-quality (LQ) images acquired within 8s and 2s, respectively. The HQ and LQ images were obtained by averaging 16 and 4 frames, respectively, and each has a spatial resolution of 512 $\times$ 512 pixels for a 450 $\times$ 450  $\mu$ m<sup>2</sup> tile scan which is approximately equal to 0.88  $\mu$ m/pixel.

### 3.2. Workflow

As mentioned before, we compared various denoising methods; the phase retrieval via GS, the median filter method, the pre-trained deep network, DnCNN, the N2N network, the MIRNet network, and our incSRCNN network. In the GS algorithm, each modality of the nonlinear multimodal imaging is processed independently. Since this algorithm depends greatly on knowing the source object, a Gaussian estimation is incorporated into the algorithm. Similarly, the MF method is applied directly to each channel of the MM images for a 3  $\times$  3 kernel size. For the DnCNN network, the pre-trained network was loaded and employed separately on each of the modalities of the nonlinear multimodal images to predict high-quality images. In the case of the N2N, the MIRNet, and our proposed network, data augmentation is applied. Before data augmentation, one image was left aside for testing, and nine were split into 7 for training and 2 for validation. Various techniques can be considered for data augmentation; however, we used rotation, blurring, and Poisson noise for our medical images. In the analysis, we first created artificial LQ images by generating Poisson noise from the HQ images. The experiment and the artificial LQ images are rotated by 90°, 180°, and 270°, and the experiment LQ images were blurred using a Gaussian filter. The total number of images equals 63 for the training part and 18 for the validation. We simultaneously applied data augmentation for both HQ and LQ images. In addition, each image was split into 16 patches. Consequently, the total patch images in the training and validation sets are 1008 and 288 patch images for each channel, respectively. Since each modality of the nonlinear multimodal imaging techniques measures specific molecular contributions, for instance, CARS modality explores the molecular distribution of proteins and lipids, SHG modality highlights collagen distribution in the sample, and TPEF modality identifies specific molecules like keratin and NADP(H), we considered these channels as independent images. Accordingly, the total number of patch images equals 3024 and 864 for the training and validation sets, respectively. However, only the CARS channel is used to train the network modality with 1008 and 288 images for training and validation sets, respectively. Since the CARS channel includes more structures while the background is more prominent in both the TPEF and the SHG modalities.

## 4. Results

Our analysis was split into two sections; first, we created artificial low-quality (LQ) images by generating Poisson noise from high-quality images (HQ). These artificial LQ images are created intentionally of lower quality as our experimental low-quality images are to be used subsequently in the training of the N2N, the MIRNet, and the incSRCNN deep learning networks. Therefore, the trained deep learning networks can generalize and cover other measurements with a different

setup and lower quality. We then evaluated the GS algorithm, the MF method, the pre-trained DnCNN, the trained N2N, the trained MIRNet, and the trained incSRCNN networks on these artificial LQ images. We generated Poisson noise from HQ images, which results in an average PSNR decrease from 19.7 to 16.4. Finally, we tested all these methods on the experimental LQ image and compared their performances. However, image reconstruction evaluation is a tricky task, particularly for medical images, and as far as the authors know, no image metric is (always) recommended. Therefore, we used a panel of image metrics: the peak signal-to-noise ratio (PSNR), the structural similarity index measure (SSIM), the image correlation coefficient (ICC) [43], and the mean absolute error (MAE). Briefly, the PSNR is a widely used metric for measuring the quality of an image which compares the original image  $x$  to the reconstructed one  $\hat{x}$  by calculating the ratio of the peak signal to the noise and it can be formulated mathematically as follows

$$PSNR(x, \hat{x}) = 10 \times \log_{10} \left( \frac{Max_x}{\sqrt{\frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \|x(i, j) - \hat{x}(i, j)\|^2}} \right), \quad (6)$$

where  $I$  represents the number of rows of pixels of the image and  $J$  is the number of columns of pixels of the image. SSIM, on the other hand, is used to quantify the similarity of a reconstructed image to its original one. It compares the luminance, contrast, and structure of the two images, and it is given by the following equation

$$SSIM(x, \hat{x}) = \frac{(2\bar{x}\bar{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(x^2 + \bar{x}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)}, \quad (7)$$

where  $\bar{x}$  is the mean of the original image,  $\bar{\hat{x}}$  is the mean of the reconstructed image,  $\sigma_x$  is the standard deviation of the original image,  $\sigma_{\hat{x}}$  is the standard deviation of the reconstructed image,  $C_1$  and  $C_2$  are two constants, and  $\sigma_{x\hat{x}}$  represents the covariance of the original image  $x$  and the reconstructed one  $\hat{x}$ . ICC measures how well two images are correlated, and it is calculated as follows

$$ICC(x, \hat{x}) = \sigma_{x\hat{x}} / \sigma_x \sigma_{\hat{x}}, \quad (8)$$

and finally, MAE measures the average magnitude of the errors between the constructed values and the original one and it is given by the following formula

$$MAE(x, \hat{x}) = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J |x(i, j) - \hat{x}(i, j)|. \quad (9)$$

In addition, we visualized the residual images and the histogram of the residual images. Moreover, the time for reconstructing one channel by these methods was illustrated in Table S1 in [Supplement 1](#).

First, the GS algorithm is implemented independently on the three channels that form the MM LQ images. The GS algorithm requires an approximation of the source beam and the LQ image as input. Therefore, the source beam is represented by Gaussian approximation (its illustration is shown as **X** in Fig. 2). A detailed explanation of the GS algorithm is discussed in the method section. The number of iterations that the algorithm carries on is 50000, and the code was built using Matlab 2020b (The MathWorks, Natick, MA).

The GS reconstruction of the artificial LQ image, displayed in Fig. 4(c-1), generally preserves the structure but includes dark regions resulting from the Gaussian estimation. Furthermore, the overall similarity between the HQ and reconstructed images was significantly low. Since the CARS channel has a more complex structure than the TPEF and SHG channels where the background is more prominent, its reconstruction compared to the other two differs dramatically with an increase from 0.27 to 0.39 in the CARS channel to an increase from 0.17 and 0.28 to 0.53

and 0.56 in the TPEF and SHG channels, respectively. In addition, although the noise level for the three channels decreases, only a slight improvement in the PSNR from 14.8 to 14.9 is shown for the CARS channel. However, the increase in the PSNR reached 21.4 and 21.7 from 16.1 and 18.4 for TPEF and SHG channels, respectively (refer to Table 1 for more details). Moreover, the SSIM is improved for the three channels. Similar to the PSNR, an increase in the SSIM value is deduced in the three channels' reconstruction. For instance, the SSIM for the CARS, TPEF, and SHG reconstructions increase from 0.27, 0.17, and 0.28 to 0.39, 0.53, and 0.56, respectively. Furthermore, the average ICC and MAE of the reconstructed image are equal to 0.86 and 0.09, respectively, while their values for the artificial image were 0.68 and 0.12. For more insights about the GS reconstructions, the residual images of the HQ and the reconstruction images are visualized in Fig. S1 in [Supplement 1](#) and compared to the residual images of the HQ and the artificial LQ images. Although the GS reconstruction was able to reconstruct some parts of the image compared with the artificial LQ case, various regions are still not well reconstructed. In addition, we illustrated in [Supplement 1](#) Fig. S2 the histogram of the residual of the reconstructed image and compared it with the residual of the artificial LQ image. Although the PSNR, SSIM, ICC, and MAE showed improved values and the histogram of the residual images showed fewer variations, the reconstructions were poor and revealed darker regions.

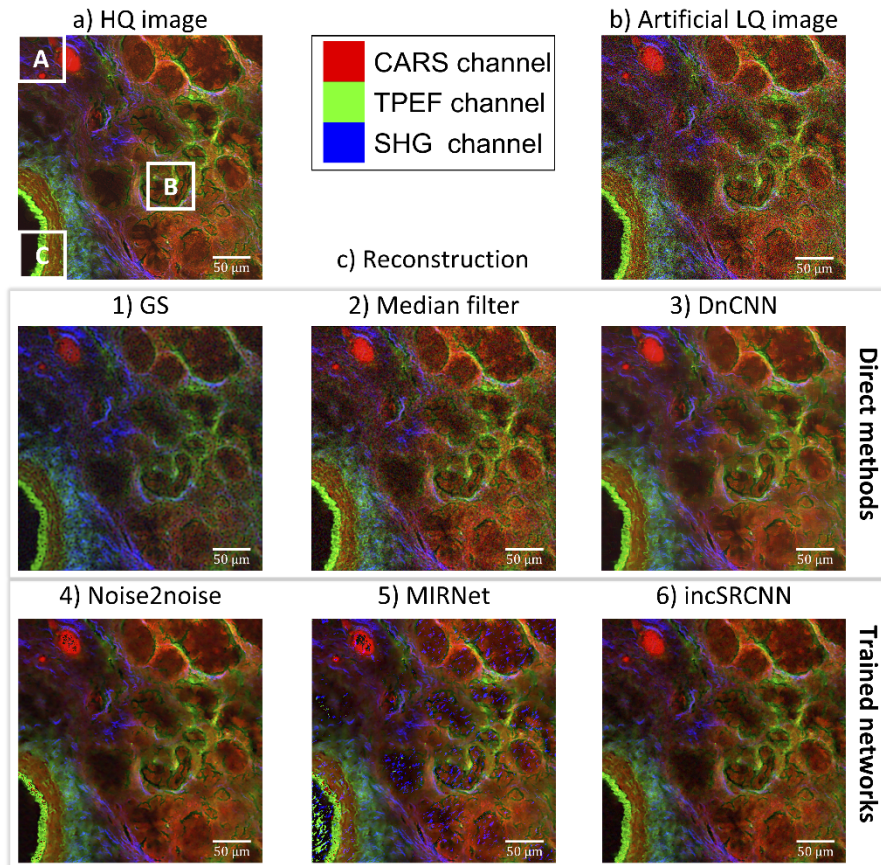
**Table 1. The PSNR, the SSIM, the ICC, and the MAE between the HQ image and the artificial LQ image and between the HQ image and the reconstructed images using the GS algorithm, the MF method, the DnCNN, N2N, MIRNet, and incSRCNN networks**

Image	Metric	Artificial LQ	GS	Median Filter	DnCNN	N2N	MIRNet	incSRCNN
CARS channel	PSNR	14.8	14.9	20.1	<b>23.0</b>	20.6	20.1	20.5
	SSIM	0.27	0.39	0.43	<b>0.59</b>	0.56	0.56	0.54
	ICC	0.65	0.83	0.83	<b>0.90</b>	0.88	0.85	0.89
	MAE	0.14	0.15	0.08	<b>0.05</b>	0.07	0.07	0.07
TPEF channel	PSNR	16.1	21.4	22.0	<b>26.2</b>	22.2	19.9	22.4
	SSIM	0.17	0.53	0.37	<b>0.64</b>	0.59	0.57	0.55
	ICC	0.65	0.89	0.86	<b>0.94</b>	0.91	0.80	0.93
	MAE	0.12	0.06	0.06	<b>0.03</b>	0.06	0.06	0.06
SHG channel	PSNR	18.4	21.7	22.0	<b>25.4</b>	22.1	13.9	21.9
	SSIM	0.28	0.56	0.42	<b>0.77</b>	0.65	0.37	0.57
	ICC	0.75	0.85	0.85	<b>0.93</b>	0.90	0.46	0.91
	MAE	0.09	0.06	0.06	<b>0.03</b>	0.05	0.10	0.06
MM image	PSNR	16.4	19.3	21.3	<b>24.9</b>	21.6	17.9	21.6
	SSIM	0.24	0.49	0.41	<b>0.67</b>	0.60	0.50	0.56
	ICC	0.68	0.86	0.85	<b>0.92</b>	0.90	0.70	0.91
	MAE	0.12	0.09	0.07	<b>0.04</b>	0.06	0.08	0.06

Next, we applied the median filter as a second standard method. The MF reconstruction of the artificial LQ image, displayed in Fig. 4(c-2), preserves the overall structure of the image. Although the PSNR, the SSIM, the ICC, and the MAE metrics showed improved values in the MF reconstruction (refer to Table 1 for more details), the MF method could not completely remove the noise.

Afterward, we applied a pre-trained network, the DnCNN, to predict the reconstruction of the artificial MM images. The DnCNN was implemented in Matlab 2020b (The MathWorks, Natick, MA). Similar to the GS algorithm, the DnCNN network was used independently on each of the three modalities. The reconstruction of the artificial LQ image is shown in Fig. 4(c-3). The spatial structures in the image are preserved, and the noise level is reduced. Furthermore, the PSNR has

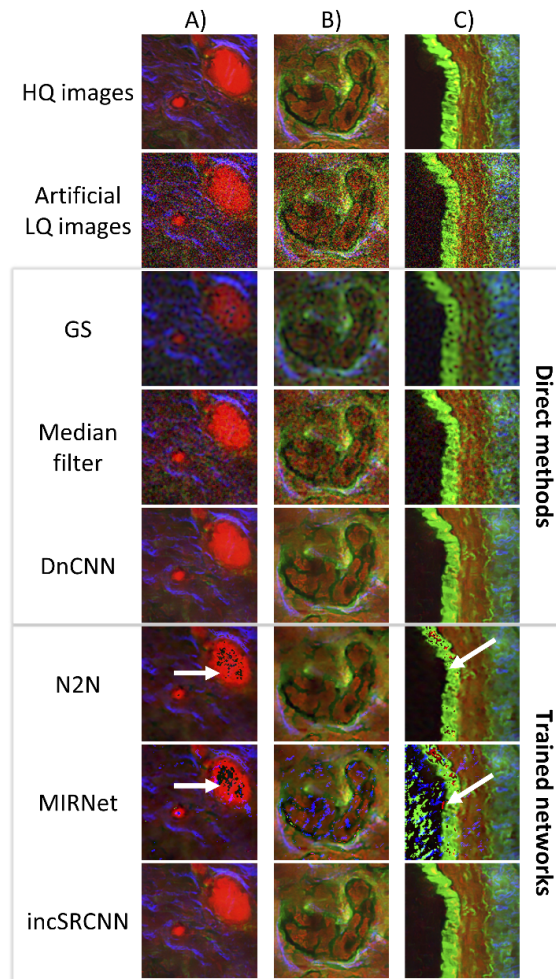




**Fig. 4.** The artificial LQ image with corresponding reconstructions using direct methods via the GS algorithm, the MF method, and the DnCNN, and the results using trained networks via the Noise2Noise (N2N), the MIRNet, and the incSRCNN networks. The experimental HQ and artificial LQ images are displayed in a) and b). The reconstructions of the artificial LQ image using the GS algorithm, the MF method, the DnCNN network, the N2N network, the MIRNet network, and the incSRCNN network are shown in part c subpanel 1,2,3,4,5,6, respectively. At first glance, the DnCNN network represents the HQ image better. On the other hand, the trained N2N and MIRNet networks show inefficiency in some regions due to the lack of data. Moreover, the proposed incSRCNN network preserves detailed structures compared to the smooth region produced by the DnCNN network. Still, some artifacts were produced, resulting from the small data size used to train the network. All the MM images represent CARS, TPEF, and SHG modalities as the red, green, and blue channels, respectively.

increased to 23.03, 26.2, and 25.4 for the CARS, TPEF, and SHG channels from 14.8, 16.1, and 18.4, respectively. In addition, the SSIM has significantly increased from 0.27, 0.17, and 0.28 to 0.59, 0.64, and 0.77 for the CARS, TPEF, and SHG channels, respectively. Consequently, the overall PSNR and SSIM improved from 16.4 and 0.24 to 24.9 and 0.67, respectively. Furthermore, the average ICC and MAE of the reconstructed image are equal to 0.92, and 0.04 while their values for the artificial image were 0.68, and 0.12. Figure 5 shows three regions of interest (ROIs) for all reconstruction algorithms. The colors in the DnCNN reconstruction are well conserved, and the noise level in the reconstruction was reduced significantly. However, smoothed structures are displayed in these ROIs, which is critical for biomedical applications because

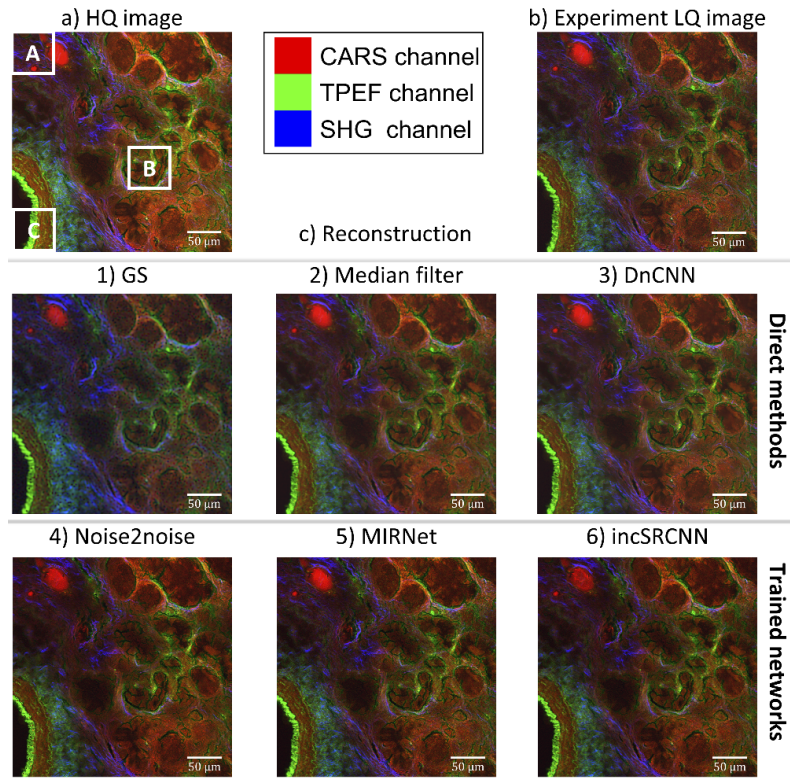




**Fig. 5.** Region of interests (ROIs) of the HQ image, the artificial LQ image, and its reconstructions using the GS algorithm, the MF method, the DnCNN, the N2N, the MIRNet, and the incSRCNN networks. The GS algorithm produces blurry images with dark spots/regions. Moreover, the MF method is not able to remove completely the noise. In the DnCNN reconstruction, some fine structures were lost while these structures were preserved using the incSRCNN network. However, some black dots were produced as artifacts using the trained N2N, MIRNet, and incSRCNN networks, which resulted from the small data size used to train the network. All the MM images represent CARS, TPEF, and SHG modalities as the red, green, and blue channels, respectively.

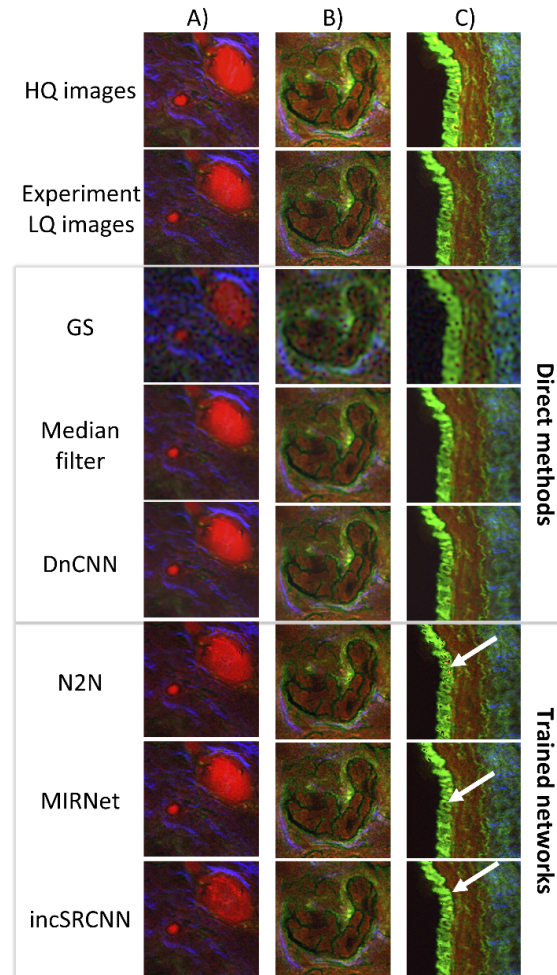
some important information may be compromised and lost, affecting the diagnosis of tissue abnormalities and diseases. In addition, the residual images per channel between the HQ image and the reconstructed one are shown in [Supplement 1 Fig. S1](#). In this figure, most of the values across the image are zero, which means that the DnCNN was able to reconstruct the exact value of the high-quality image.

Moreover, the histogram of the residual images between the HQ image and the DnCNN reconstruction in [Supplement 1 Fig. S2](#) significantly reduces the values compared with the artificial LQ case. Therefore, the DnCNN reconstructed a good representation of the HQ image successfully.



**Fig. 6.** The experimental LQ image with corresponding reconstructions using direct methods via the GS algorithm, the MF method, and the DnCNN and trained networks via the Noise2Noise (N2N), the MIRNet, and the incSRCNN networks. The experimental HQ and LQ images are displayed in a) and b), respectively. The reconstruction of the experiment LQ image using the GS algorithm, the MF method, the DnCNN network, the N2N network, the MIRNet network, and the incSRCNN network is shown in part c / subpanel 1,2,3,4,5,6, respectively. At first glance, the DnCNN network better represents the HQ image. However, the N2N, MIRNet, and incSRCNN reconstructions preserve detailed structures while the DnCNN reconstruction displays smoothed structures. All the MM images represent CARS, TPEF, and SHG modalities as the red, green, and blue channels, respectively.

Finally, we evaluated our proposed network (incSRCNN) on the same artificial LQ images and compare it with two deep learning networks via the Noise2Noise and MIRNet networks that were trained with the same augmented MM images. The detailed architecture was described in the method section. The training of the network was performed by minimizing the mean absolute error (MAE)-based loss between the HQ images and the output of the incSRCNN network. The Adam algorithm was used for the optimization with a learning rate of  $3e^{-4}$ . A total of 1008 and 288 coupled HQ and LQ images were used for the training and the validation, respectively; refer to [Supplement 1 Table S2](#) for more details. All computations were done using Google Colab. The total number of parameters to be trained is 20,481 (refer to Fig. S7 and Fig. S8 for more details about the architecture and parameters of the incSRCNN network). The training and prediction time for our architecture is around 10 minutes and 7 seconds, respectively. However, the N2N and MIRNet networks require around 48 minutes and 3 hours in the training phase and 33 and 68 seconds for the reconstruction of the image (refer to Table S1 for more details). We assessed different cases to train the network; three independent incSRCNN on each modality, one



**Fig. 7.** Region of interest (ROIs) of the HQ image, the experimental LQ image, and its constructions using the GS algorithm, the MF method, the DnCNN, the N2N, the MIRNet, and the incSRCNN networks. The GS algorithm produces a dark region due to the Gaussian estimation. Moreover, the MF reconstruction could not remove the noise completely. In the DnCNN reconstruction, some fine structures are lost while preserved using the incSRCNN network. However, some dark dots were produced, resulting from the lack of data used to train the 3 networks; N2N, MIRNet, and incSRCNN. All the MM images represent CARS, TPEF, and SHG modalities as the red, green, and blue channels, respectively.

incSRCNN comprising all channels as separate data, and one incSRCNN that includes only the CARS channel. We found out that training with only the CARS channel produces better results. The training time of this network is approximately 10 minutes compared to 1 hour in the first and second cases.

The incSRCNN reconstruction of the artificial LQ image is shown in Fig. 4(c-6). While the N2N and MIRNet reconstructions are shown in the same Fig. 4(c-4 and 5, respectively). The spatial structures and the color in the images are preserved, and the noise level is reduced. Furthermore, the PSNR has increased to 20.5, 22.4, and 21.9 for the CARS, TPEF, and SHG channels from 14.8, 16.1, and 18.4, respectively. In addition, the SSIM has significantly increased

from 0.27, 0.17, and 0.28 to 0.54, 0.55, and 0.57 for the CARS, TPEF, and SHG channels, respectively. Consequently, the overall PSNR and SSIM improved from 16.4 and 0.24 to 21.6 and 0.56, respectively. Furthermore, the average ICC and MAE of the reconstructed image are equal to 0.91, and 0.06, while their values for the artificial image were 0.68, and 0.12. Figure 5 shows three regions of interest (ROIs) of all reconstructions. The colors in the incSRCNN reconstruction are well conserved, and the noise level in the reconstruction was reduced significantly. In addition, the residual images per channel between the HQ image and the reconstructed one are shown in Supplement 1 Fig. S1. In this figure, a significant reduction of the STD values in the residual images from 0.2, 0.2, and 0.1 for CARS, TPEF, and SHG of the artificial LQ case, respectively to 0.08, 0.06, and 0.06 for CARS, TPEF, and SHG of the incSRCNN case, respectively.

Moreover, the histogram of the residual images between the HQ image and the incSRCNN reconstruction in Supplement 1 Fig. S2 significantly reduces the values compared with the artificial LQ case. Therefore, the incSRCNN reconstructed a good representation of the HQ image successfully. Compared to the other more complex deep learning networks N2N and MIRNet, all three networks produced black spots which resulted from the lack of data in the training procedure.

The next step is to assess the six methods on the experimental LQ MM image. First, we evaluated the GS algorithm on the experimental LQ image with the exact source estimation used for the artificial LQ image.

In the experimental LQ image reconstruction using the GS algorithm, the PSNR for the TPEF channel increased from 20.0 to 20.1. However, the PSNR decreased from 19.0 and 20.1 to 14.9 and 20.1 in the CARS and SHG channels, respectively. In addition, the SSIM improved for only the TPEF channel but worsened in the CARS and the SHG channels. All characteristics are given in Table 2. Furthermore, the average ICC showed an improved correlation value from 0.78 to 0.80, but the average MAE showed an increased value. It is worth noticing that the worsened values are mainly related to the CARS channel reconstruction since the other channels presented acceptable results. Moreover, the GS reconstruction of the experimental LQ image does not differ from the artificial LQ reconstruction, which can be deduced from the residual images and the residual histogram in Supplement 1 Fig. S3 and Fig. S4, respectively. The reason is that the algorithm converged to a local minimum and could not improve more.

Then, in the experimental LQ image reconstruction using the MF method, the PSNR for the CARS reconstruction decreased from 19.0 to 18.8. In addition, the SSIM of the CARS reconstruction also dropped to 0.54 from 0.56. We could conclude that the experimental LQ image is already in a high-quality condition that even a filter-based method could not improve the quality of the CARS channel.

Afterward, we tested the performance of the DnCNN in the experimental LQ image. In Fig. 6(c-3), we showed the reconstruction using the DnCNN network, where the spatial structure in the image is preserved, and the noise level is slightly reduced. In Table 2, we compared the PSNR, SSIM, ICC, and MAE between the DnCNN reconstructions and the HQ image with the PSNR, SSIM, ICC, and MAE between the experiment LQ image and the HQ image. Compared to the experiment LQ image, we deduced a slight improvement in the PSNR, SSIM, ICC, and MAE values per channel and overall, when using the DnCNN network. In Fig. 7, three ROIs showed a reduction in the noise level. Like the artificial case, smoothed regions were produced, which might cause the removal of important features that are highly sensitive in the diagnosis of diseases and abnormalities. In addition, we compared the residual images per channel of the DnCNN reconstructions with the residual images of the experiment LQ image in Supplement 1 Fig. S3. In this figure, almost similar residual values to the experimental LQ case can be detected. Furthermore, we visualized the histogram of the residual images of the DnCNN reconstruction and the experiment LQ images in Supplement 1 Fig. S4.



**Table 2. The PSNR, the SSIM, the ICC, and the MAE between the HQ and the experimental LQ images and between the HQ and the reconstructed images using the GS algorithm, the MF method, the DnCNN, N2N, MIRNet, and incSRCNN networks**

Image	Metric	LQ	GS	Median Filter	DnCNN	N2N	MIRNet	incSRCNN
CARS channel	PSNR	19.0	14.9	18.8	<b>19.2</b>	17.7	17.8	17.2
	SSIM	0.56	0.38	0.54	<b>0.60</b>	0.54	0.54	0.53
	ICC	0.86	0.82	0.87	<b>0.88</b>	0.85	0.86	0.86
	MAE	0.09	0.15	0.09	<b>0.08</b>	0.11	0.11	0.11
TPEF channel	PSNR	20.0	20.1	20.6	<b>20.8</b>	18.5	18.9	18.6
	SSIM	0.46	0.54	0.60	<b>0.62</b>	0.51	0.44	0.47
	ICC	0.76	0.80	<b>0.82</b>	<b>0.82</b>	0.79	0.77	0.79
	MAE	0.06	0.06	0.06	<b>0.05</b>	0.09	0.08	0.09
SHG channel	PSNR	20.1	20.1	20.7	<b>20.7</b>	19.1	19.7	19.6
	SSIM	0.63	0.54	<b>0.68</b>	<b>0.68</b>	0.60	0.62	0.60
	ICC	0.73	0.77	<b>0.78</b>	0.77	0.72	0.74	0.76
	MAE	0.05	0.07	<b>0.04</b>	0.05	0.06	0.06	0.06
MM image	PSNR	19.7	18.3	20.0	<b>20.2</b>	18.4	18.8	18.4
	SSIM	0.55	0.49	0.60	<b>0.64</b>	0.55	0.53	0.54
	ICC	0.78	0.80	<b>0.82</b>	<b>0.82</b>	0.79	0.79	0.80
	MAE	0.07	0.09	<b>0.06</b>	<b>0.06</b>	0.09	0.08	0.09

Finally, we tested the performance of the incSRCNN in the experimental LQ image and compare it with the N2N and MIRNet networks. In Fig. 6(c-4, 5, 6), we showed the reconstruction using the N2N, the MIRNet, and our proposed network, where the spatial structures and the color in the image are preserved; however, the noise level is slightly reduced. In addition, we compared in Table 2 the PSNR, SSIM, ICC, and MAE between the three trained network reconstructions and the HQ image with the PSNR, SSIM, ICC, and MAE between the experiment LQ image and the HQ image. Compared to the experiment LQ image, although the average ICC improved, the PSNR and SSIM values per channel overall decreased. In Fig. 7, our proposed network preserves the color and spatial structures. However, the decrease mentioned above might result from the small data size that the network failed to estimate the values in some areas indicated by the arrow in the figure. However, we continued to assess this matter by checking the intensity values across an arbitrary region and evaluating the incSRCNN network on other noisy experimental LQ data. These noisy LQ data were derived by generating Poisson noise from the experimental LQ images. The results are illustrated in Supplement 1 Fig. S5 and Fig. S6. Figure S5 shows the intensity values for the HQ, LQ, and reconstructed images across the specified region in the image on the top left of the figure. The intensity values in GS differ totally from those in the LQ image, while the deep learning methods maintained a similar trend. However, the smoothed nature of the DnCNN reconstruction can be reflected by showing fewer details than in the incSRCNN reconstruction. Besides, the incSRCNN reconstructions for both noisy experimental LQ images illustrated in Figure S6 showed improvements in terms of PSNR, SSIM, ICC, and MAE values. In addition, we compared the residual images per channel of the incSRCNN reconstructions with the residual images of the experiment LQ image in Supplement 1 Fig. S3. In this figure, almost similar residual values to the experimental LQ case can be detected. Furthermore, we visualized the histogram of the residual images of the DnCNN reconstruction and the experiment LQ images in Supplement 1 Fig. S4. We previously discussed the performance of each method compared to the artificial and experimental LQ images. The GS reconstruction shows similar but poor performance for both artificial and experimental LQ images. The GS reconstructions include



dark regions, and the algorithm showed limited abilities even in noiseless settings. In addition, it seems that the optimization algorithm of the GS method converges to a local minimum that causes poor reconstructions. However, the DnCNN and the incSRCNN reconstructions preserved the colors and detailed structures. Both networks performed well in the artificial LQ case, but the DnCNN produced smoothed regions critical for medical applications. Our proposed network consists of a simple architecture that only uses the CARS channels and predicts the other two channels. Similar to the artificial case, the DnCNN and incSRCNN networks performed better. These two networks preserve the color and the spatial structures of the image. However, the DnCNN network produced smoothed region, which is a drawback compared to our proposed network that shows a slight reduction in the noise due to the lack of data that the network could not train some regions. We additionally trained two other networks with deeper layers; the N2N and MIRNet networks. We showed that the lack of data affects the reconstruction also in these two networks.

## 5. Generalizability

The results explained previously involve only one image position where all the methods were either applied directly on this image or trained using the remaining 9 images and then using the trained networks to reconstruct this particular image. Therefore, in this section, we want to evaluate the PSNR, SSIM, ICC, and MAE values in two different ways: patch-wise analysis and cross-validation analysis. Both methods are utilized to investigate the variability of the reconstructions within an image (patch-wise analysis) and between images (cross-validation). In the patch-wise analysis, one single testing image of size  $512 \times 512$  was used for validation. In this method, the PSNR, SSIM, ICC, and MAE are not evaluated on the whole reconstruction from both the direct methods and the trained networks, but the markers are calculated for 16 patches of size  $128 \times 128$ . This means that PSNR, SSIM, ICC, and MAE are calculated per patch resulting in 16 values per metric and then the average and standard deviation of these metrics are computed and visualized in Table 3 and Table 4. The second analysis is the cross-validation analysis and one MM image was left out within the cross-validation loop. In this study, a total of 10 MM images of size  $512 \times 512$  were predicted and each of these images was reconstructed using the direct methods: the GS algorithm, the MF method, and the DnCNN network. Then, the PSNR, SSIM, ICC, and MAE per reconstruction were calculated, and finally, the average and the standard deviation of these metrics were computed. While for training the N2N, MIRNet, and incSRCNN networks, 10 networks were trained in which one image is left for testing purposes and the remaining 9 images were used for training of the N2N, MIRnet, and incSRCNN network. In the end, the aforementioned metrics are calculated on each testing image, and the average and standard deviation of these metrics is computed and visualized in Table 5 and Table 6.

First, the results of the patch-wise analysis for both artificial and experimental LQ images are summarized in Table 3 and Table 4, respectively. The DnCNN in terms of metrics showed a higher improvement for the artificial and experimental reconstructions. While the incSRCNN in the experimental reconstruction showed less variation. These results are consistent with the overall calculation for both artificial and experimental LQ reconstructions.

For the second method, the results of the cross-validation analysis for both artificial and experimental LQ images are summarized in Table 3 and Table 4, respectively. We used in this part the leave one MM image out cross-validation. For the direct methods, the reconstruction is implemented directly on each channel for the 10 MM images, and the average of the evaluation metrics is calculated. However, for the trained networks, the analysis consists of leaving one image for testing purposes and using the remaining 9 images to train the N2N, MIRNet, and incSRCNN networks. Afterward, these trained networks were used to reconstruct the testing MM image. Furthermore, the evaluation metrics were constructed and the average of the 10 cases is calculated and summarized in the tables below for the artificial and the experimental LQ

**Table 3. The average PSNR, SSIM, ICC, and MAE between the HQ and the artificial LQ images and between the HQ and the reconstructed images using the GS algorithm, the MF method, the DnCNN, N2N, MIRNet, and incSRCNN networks for the patch wise analysis**

Image	Metric	Artificial LQ	GS	Median Filter	DnCNN	N2N	MIRNet	incSRCNN
CARS channel	PSNR	14.8 ± <b>0.68</b>	15.3 ± 1.96	20.3 ± 1.21	<b>23.4</b> ± 2.09	20.9 ± 1.61	20.5 ± 2.01	20.7 ± 1.21
	SSIM	0.27 ± 0.08	0.39 ± 0.04	0.43 ± 0.06	<b>0.59</b> ± <b>0.03</b>	0.56 ± <b>0.03</b>	0.56 ± <b>0.03</b>	0.54 ± <b>0.03</b>
	ICC	0.59 ± <b>0.01</b>	0.78 ± 0.03	0.77 ± <b>0.01</b>	<b>0.86</b> ± <b>0.01</b>	0.85 ± <b>0.01</b>	0.82 ± <b>0.01</b>	0.85 ± <b>0.01</b>
	MAE	0.14 ± 0.09	0.15 ± 0.07	0.08 ± 0.08	<b>0.05</b> ± 0.05	0.07 ± <b>0.04</b>	0.07 ± <b>0.04</b>	0.07 ± 0.05
TPEF channel	PSNR	16.2 ± <b>1.02</b>	22.1 ± 2.67	22.1 ± 1.13	<b>26.8</b> ± 2.37	22.7 ± 2.00	21.8 ± 3.29	22.5 ± 1.18
	SSIM	0.17 ± 0.04	0.53 ± 0.04	0.37 ± 0.04	<b>0.64</b> ± 0.04	0.59 ± 0.04	0.57 ± 0.05	0.55 ± <b>0.02</b>
	ICC	0.54 ± <b>0.01</b>	0.80 ± 0.02	0.74 ± <b>0.01</b>	<b>0.88</b> ± <b>0.01</b>	0.87 ± <b>0.01</b>	0.79 ± 0.02	0.86 ± <b>0.01</b>
	MAE	0.12 ± 0.13	0.06 ± 0.11	0.06 ± 0.13	<b>0.03</b> ± <b>0.07</b>	0.06 ± <b>0.07</b>	0.06 ± 0.14	0.06 ± 0.09
SHG channel	PSNR	18.5 ± <b>1.08</b>	22.3 ± 2.33	22.2 ± 1.50	<b>26.1</b> ± 2.43	22.6 ± 1.95	14.1 ± 1.62	22.2 ± 1.72
	SSIM	0.28 ± 0.09	0.56 ± 0.04	0.42 ± 0.07	<b>0.77</b> ± 0.04	0.65 ± 0.04	0.38 ± 0.09	0.57 ± <b>0.02</b>
	ICC	0.70 ± <b>0.01</b>	0.81 ± 0.02	0.80 ± <b>0.01</b>	<b>0.91</b> ± <b>0.01</b>	0.88 ± <b>0.01</b>	0.40 ± <b>0.01</b>	0.88 ± <b>0.01</b>
	MAE	0.09 ± 0.09	0.06 ± 0.06	0.06 ± 0.09	<b>0.03</b> ± <b>0.03</b>	0.05 ± 0.04	0.10 ± 0.20	0.06 ± 0.05
MM image	PSNR	16.5 ± <b>0.92</b>	19.9 ± 2.32	21.5 ± 1.28	<b>25.4</b> ± 2.30	22.1 ± 1.85	18.8 ± 2.31	21.8 ± 1.37
	SSIM	0.24 ± 0.07	0.49 ± 0.04	0.41 ± 0.06	<b>0.66</b> ± 0.04	0.60 ± 0.04	0.50 ± 0.06	0.56 ± <b>0.02</b>
	ICC	0.61 ± <b>0.01</b>	0.80 ± 0.02	0.77 ± <b>0.01</b>	<b>0.88</b> ± <b>0.01</b>	0.87 ± <b>0.01</b>	0.67 ± <b>0.01</b>	0.86 ± <b>0.01</b>
	MAE	0.12 ± 0.11	0.09 ± 0.08	0.07 ± 0.10	<b>0.04</b> ± <b>0.05</b>	0.06 ± <b>0.05</b>	0.08 ± 0.13	0.06 ± 0.06

**Table 4. The average PSNR, SSIM, ICC, and MAE between the HQ and the experimental LQ images and between the HQ and the reconstructed images using the GS algorithm, the MF method, the DnCNN, N2N, MIRNet, and incSRCNN networks for the patch wise analysis**

Image	Metric	LQ	GS	Median Filter	DnCNN	N2N	MIRNet	incSRCNN
CARS channel	PSNR	19.4 ± 1.88	15.3 ± 2.04	19.2 ± 2.15	<b>19.6</b> ± 2.02	18.0 ± <b>1.53</b>	18.0 ± 1.50	17.4 ± 1.58
	SSIM	0.56 ± 0.19	0.38 ± <b>0.06</b>	0.54 ± 0.12	<b>0.60</b> ± 0.17	0.54 ± 0.19	0.54 ± 0.19	0.53 ± 0.19
	ICC	0.79 ± <b>0.02</b>	0.75 ± 0.03	<b>0.82</b> ± <b>0.02</b>	0.81 ± <b>0.02</b>	0.78 ± <b>0.02</b>	0.78 ± <b>0.02</b>	0.79 ± <b>0.02</b>
	MAE	0.09 ± 0.16	0.15 ± <b>0.12</b>	0.09 ± <b>0.12</b>	<b>0.08</b> ± 0.14	0.11 ± 0.16	0.11 ± 0.16	0.11 ± 0.16
TPEF channel	PSNR	22.0 ± 3.63	22.1 ± 3.88	22.9 ± 4.02	<b>23.4</b> ± 4.19	19.6 ± 2.89	20.2 ± 3.05	19.7 ± <b>2.92</b>
	SSIM	0.46 ± <b>0.10</b>	0.54 ± 0.11	0.60 ± 0.13	<b>0.62</b> ± 0.13	0.51 ± 0.11	0.44 ± <b>0.10</b>	0.47 ± <b>0.10</b>
	ICC	0.72 ± 0.03	0.77 ± 0.03	0.81 ± 0.03	<b>0.83</b> ± 0.03	0.78 ± 0.03	0.73 ± 0.03	0.76 ± 0.03
	MAE	0.06 ± 0.17	0.06 ± <b>0.16</b>	0.06 ± <b>0.16</b>	<b>0.05</b> ± <b>0.16</b>	0.09 ± 0.18	0.08 ± 0.17	0.09 ± 0.17
SHG channel	PSNR	22.3 ± 4.67	21.1 ± <b>3.01</b>	22.5 ± 4.29	<b>23.1</b> ± 4.95	20.6 ± 3.70	21.5 ± 4.24	21.1 ± 3.74
	SSIM	0.63 ± 0.19	0.54 ± <b>0.12</b>	<b>0.68</b> ± 0.17	<b>0.68</b> ± 0.18	0.60 ± 0.16	0.62 ± 0.18	0.60 ± 0.16
	ICC	0.74 ± 0.03	0.75 ± <b>0.02</b>	<b>0.79</b> ± <b>0.02</b>	0.78 ± 0.03	0.74 ± 0.03	0.75 ± 0.03	0.77 ± 0.03
	MAE	0.05 ± 0.16	0.07 ± <b>0.12</b>	<b>0.04</b> ± 0.14	0.05 ± 0.16	0.06 ± 0.14	0.06 ± 0.16	0.06 ± 0.15
MM image	PSNR	21.2 ± 3.39	19.5 ± 2.98	21.5 ± 3.49	<b>22.0</b> ± 3.72	19.4 ± <b>2.71</b>	19.9 ± 2.93	19.4 ± 2.75
	SSIM	0.55 ± 0.16	0.49 ± <b>0.10</b>	0.60 ± 0.14	<b>0.64</b> ± 0.16	0.55 ± 0.15	0.53 ± 0.16	0.54 ± 0.15
	ICC	0.75 ± 0.03	0.76 ± 0.03	<b>0.81</b> ± 0.03	<b>0.81</b> ± <b>0.02</b>	0.76 ± <b>0.02</b>	0.75 ± <b>0.02</b>	0.77 ± <b>0.02</b>
	MAE	0.07 ± 0.16	0.09 ± <b>0.13</b>	<b>0.06</b> ± 0.14	<b>0.06</b> ± 0.15	0.09 ± 0.16	0.08 ± 0.16	0.09 ± 0.16

**Table 5. The average PSNR, SSIM, ICC, and MAE between the HQ and the artificial LQ images and between the HQ and the reconstructed images using the GS algorithm, the MF method, the DnCNN, N2N, MIRNet, and incSRCNN networks for the cross-validation analysis**

Image	Metric	Artificial LQ	GS	Median Filter	DnCNN	N2N	MIRNet	incSRCNN
CARS channel	PSNR	14.6 ± <b>0.31</b>	14.9 ± 1.31	20.2 ± 0.54	<b>23.6</b> ± 0.95	21.8 ± 0.90	21.7 ± 1.31	21.7 ± 0.89
	SSIM	0.23 ± 0.05	0.37 ± 0.04	0.39 ± 0.04	<b>0.55</b> ± 0.04	0.54 ± <b>0.03</b>	0.54 ± 0.04	0.52 ± <b>0.03</b>
	ICC	0.56 ± 0.07	0.76 ± 0.03	0.75 ± <b>0.00</b>	<b>0.86</b> ± 0.01	0.86 ± 0.01	0.85 ± 0.01	0.84 ± 0.01
	MAE	0.15 ± <b>0.01</b>	0.16 ± 0.07	0.08 ± 0.07	<b>0.05</b> ± 0.04	0.06 ± 0.03	0.06 ± 0.04	0.06 ± 0.04
TPEF channel	PSNR	17.6 ± <b>0.98</b>	21.9 ± 2.30	22.8 ± 1.00	<b>27.3</b> ± 2.19	22.3 ± 1.84	21.1 ± 2.73	22.8 ± 2.01
	SSIM	0.20 ± 0.07	0.56 ± <b>0.05</b>	0.41 ± 0.07	<b>0.75</b> ± 0.08	0.70 ± 0.07	0.52 ± 0.07	0.65 ± <b>0.05</b>
	ICC	0.67 ± 0.13	0.87 ± 0.01	0.84 ± 0.01	<b>0.94</b> ± 0.01	0.87 ± <b>0.00</b>	0.72 ± 0.01	0.91 ± 0.01
	MAE	0.10 ± <b>0.01</b>	0.06 ± 0.05	0.05 ± 0.09	<b>0.03</b> ± 0.03	0.06 ± 0.02	0.07 ± 0.06	0.06 ± 0.02
SHG channel	PSNR	18.7 ± <b>0.68</b>	21.2 ± 1.59	22.9 ± 0.92	<b>26.8</b> ± 1.98	23.2 ± 1.73	20.1 ± 5.01	22.7 ± 0.95
	SSIM	0.25 ± 0.05	0.52 ± 0.07	0.41 ± <b>0.04</b>	<b>0.75</b> ± <b>0.04</b>	0.66 ± 0.07	0.52 ± 0.17	0.54 ± <b>0.04</b>
	ICC	0.67 ± 0.11	0.83 ± 0.02	0.79 ± <b>0.00</b>	<b>0.91</b> ± 0.01	0.87 ± 0.01	0.72 ± 0.04	0.87 ± 0.01
	MAE	0.09 ± <b>0.01</b>	0.07 ± 0.06	0.05 ± 0.11	<b>0.03</b> ± 0.04	0.05 ± 0.07	0.07 ± 0.23	0.05 ± 0.06
MM image	PSNR	17.9 ± <b>0.66</b>	19.3 ± 1.74	22.0 ± 0.82	<b>25.9</b> ± 1.71	22.4 ± 1.49	20.9 ± 3.02	22.4 ± 1.28
	SSIM	0.23 ± 0.06	0.48 ± 0.05	0.40 ± 0.05	<b>0.68</b> ± 0.05	0.63 ± 0.06	0.53 ± 0.09	0.57 ± <b>0.04</b>
	ICC	0.63 ± 0.10	0.82 ± 0.02	0.79 ± <b>0.01</b>	<b>0.90</b> ± <b>0.01</b>	0.87 ± <b>0.01</b>	0.76 ± 0.02	0.87 ± <b>0.01</b>
	MAE	0.11 ± <b>0.01</b>	0.10 ± 0.06	0.06 ± 0.09	<b>0.04</b> ± 0.04	0.06 ± 0.04	0.07 ± 0.11	0.06 ± 0.04

**Table 6. The average PSNR, SSIM, ICC, and MAE between the HQ and the experimental LQ images and between the HQ and the reconstructed images using the GS algorithm, the MF method, the DnCNN, N2N, MIRNet, and incSRCNN networks for the cross-validation analysis**

Image	Metric	LQ	GS	Median Filter	DnCNN	N2N	MIRNet	incSRCNN
CARS channel	PSNR	18.6 ± 1.31	14.9 ± 1.36	18.9 ± 1.37	<b>19.1</b> ± 1.38	14.2 ± 1.73	12.8 ± 2.23	15.3 ± <b>0.87</b>
	SSIM	0.39 ± 0.08	0.33 ± <b>0.05</b>	0.44 ± 0.06	<b>0.47</b> ± 0.07	0.27 ± 0.13	0.17 ± 0.17	0.39 ± 0.07
	ICC	0.72 ± 0.10	0.73 ± 0.03	0.78 ± <b>0.01</b>	0.78 ± 0.01	0.52 ± 0.03	0.30 ± 0.04	<b>0.79</b> ± 0.02
	MAE	<b>0.09</b> ± <b>0.01</b>	0.16 ± 0.09	<b>0.09</b> ± 0.08	<b>0.09</b> ± 0.08	0.16 ± 0.20	0.18 ± 0.23	0.15 ± 0.07
TPEF channel	PSNR	22.5 ± 3.19	21.2 ± 2.67	23.0 ± 3.56	<b>23.4</b> ± 3.93	18.5 ± 2.44	15.5 ± 4.31	17.7 ± <b>1.66</b>
	SSIM	0.58 ± 0.11	0.60 ± <b>0.04</b>	0.72 ± 0.09	<b>0.74</b> ± 0.09	0.59 ± 0.11	0.35 ± 0.20	0.35 ± 0.16
	ICC	0.82 ± 0.06	0.83 ± 0.02	0.87 ± <b>0.01</b>	<b>0.87</b> ± <b>0.01</b>	0.74 ± 0.02	0.52 ± 0.05	0.85 ± 0.02
	MAE	0.05 ± <b>0.01</b>	0.06 ± 0.05	<b>0.04</b> ± 0.05	<b>0.04</b> ± 0.05	0.09 ± 0.07	0.12 ± 0.18	0.11 ± 0.06
SHG channel	PSNR	22.9 ± 2.63	20.3 ± <b>1.17</b>	23.4 ± 2.79	<b>23.4</b> ± 2.73	20.2 ± 2.63	15.8 ± 4.49	18.8 ± 1.53
	SSIM	0.62 ± 0.09	0.52 ± <b>0.05</b>	0.67 ± 0.09	<b>0.68</b> ± 0.08	0.59 ± 0.10	0.38 ± 0.18	0.27 ± 0.15
	ICC	0.76 ± 0.07	0.78 ± 0.02	0.80 ± <b>0.01</b>	<b>0.79</b> ± <b>0.01</b>	0.68 ± 0.02	0.43 ± 0.04	0.78 ± 0.02
	MAE	<b>0.04</b> ± <b>0.01</b>	0.08 ± 0.06	<b>0.04</b> ± 0.05	<b>0.04</b> ± 0.06	0.06 ± 0.07	0.10 ± 0.15	0.09 ± 0.05
MM image	PSNR	21.3 ± 2.38	18.8 ± 1.73	21.8 ± 2.57	<b>22.0</b> ± 2.68	17.6 ± 2.27	14.7 ± 3.68	17.3 ± <b>1.36</b>
	SSIM	0.53 ± 0.09	0.48 ± <b>0.05</b>	0.61 ± 0.08	<b>0.63</b> ± 0.08	0.48 ± 0.11	0.30 ± 0.18	0.34 ± 0.12
	ICC	0.77 ± 0.07	0.78 ± 0.02	<b>0.82</b> ± <b>0.01</b>	0.81 ± <b>0.01</b>	0.65 ± 0.02	0.42 ± 0.04	0.81 ± 0.02
	MAE	0.06 ± 0.01	0.10 ± <b>0.06</b>	<b>0.06</b> ± <b>0.06</b>	<b>0.06</b> ± <b>0.06</b>	0.10 ± 0.12	0.13 ± 0.19	0.12 ± <b>0.06</b>

images. The DnCNN in terms of metrics showed a higher improvement for the artificial and experimental reconstructions. While the incSRCNN network in the artificial and experimental reconstruction showed less variation.

## 6. Conclusion

The multimodal imaging approach (MM), which combines the CARS, the TPEF, and the SHG modalities provide information on the structure of the measured tissue and its components. However, the MM approach offers high-quality images only, when they are measured longer compared to faster MM image measurements, which results in MM images being distorted with noise and other artifacts. Therefore, image denoising techniques are helpful when fast measurements are needed or carried out. However, image denoising techniques feature the drawback that a suitable method needs to be chosen for different settings, which varies between application scenarios. In this context, we compared two classical methods; the median filter (MF) and the phase retrieval method via Gerchberg-Saxton (GS) with two deep learning approaches. The first approach is to use transfer learning via the pre-trained network namely DnCNN and the second approach is to use augmented MM images to train a deep learning network. In this context, we trained three networks; the N2N network, the MIRNet network, and our built-in network the incSRCNN. The data consists of MM images of the neck and head tissue of a mouse. First, we evaluated the GS algorithm, the MF method, the DnCNN, the N2N, the MIRNet, and the incSRCNN networks on artificial LQ images. Afterward, we tested all these methods on an experimental LQ image.

The artificial LQ image was constructed by generating Poisson noise from the HQ image. The GS algorithm of the artificial LQ image showed poor reconstruction, where dark regions are produced due to the Gaussian estimation used to describe the input beam. In addition, the MF reconstruction could not remove completely the noise. However, the DnCNN and the incSRCNN reconstructions preserve the color and the spatial structures in the image and improve the PSNR, SSIM, ICC, MAE, and STD compared to the artificial LQ image. However, the DnCNN produced smoothed region that might cause a compromise in the diagnosis of diseases and abnormalities. When comparing our incSRCNN network with the trained N2N and MIRNet networks, we concluded that the incSRCNN reconstruction is better since more black spots are produced by the MIRNet.

Afterward, we compared the performance of the six methods on the experimental LQ image. Like the artificial case, the GS algorithm showed poor performance, the MF showed good reconstruction, and the DnCNN network preserved the color and spatial structures in the images, but smoothed regions were produced. However, the incSRCNN networks maintained the color and the spatial structures in the image and did not produce smoothed areas. However, our proposed network showed a slight decrease in the PSNR, which resulted from the lack of data. In conclusion, a priori knowledge of the beam source is vital for the GS reconstruction, and the algorithm has limited recovery abilities even in a noiseless setting.

In summary, deep learning networks produced very promising results. However, the DnCNN network preserved the color and spatial structures of the image but produced smoothed regions, resulting in the loss of relevant information. However, our proposed network, the incSRCNN, consists of simple architecture, and it reconstructs the complex structures of the testing image and shows good PSNR than the other standard methods. Nevertheless, the incSRCNN network produced some artifacts represented by arrows in the zoomed figures, resulting from the lack of data used to train the network. It is worth mentioning that in all implemented methods the SSIM was around 0.6 which is quite low and this fact needs more in-depth analysis that might suggest the best evaluation metrics that can be used for denoising MM images, which we plan to investigate further as part of our future research. On the other hand, the shorter time to

reconstruct an HQ image, run on a limited CPU computer, is through the incSRCNN network. Additionally, only 0.08 second is needed for predicting a patch of the image.

**Funding.** Freistaat Thüringen (2019 FGR 0083 (Morphotox), 5575/10-9 (Digleben)); Horizon 2020 Framework Programme (101016923 (CRIMSON)); Bundesministerium für Bildung und Forschung (13GW0370E (TheraOptik), 13N15706 (LPI-BT2-FSU), 13N15710 (LPI-BT3-FSU), 13N15464 (LPI-BT1-Leibniz-IPHT)); Thueringer Universitaets- und Landesbibliothek Jena Open Access Publication Fund; German Research Foundation (512648189).

**Acknowledgments.** We acknowledge support by the German Research Foundation Projekt-Nr. 512648189 and the Open Access Publication Fund of the Thueringer Universitaets- und Landesbibliothek Jena.

**Disclosures.** The authors declare no conflicts of interest.

**Data availability.** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

**Supplemental document.** See [Supplement 1](#) for supporting content.

## References

1. A. Bhandary, G. A. Prabhu, V. Rajinikanth, K. P. Thanaraj, S. C. Satapathy, D. E. Robbins, C. Shasky, Y.-D. Zhang, J. M. R. S. Tavares, and N. S. M. Raja, "Deep-learning framework to detect lung abnormality – A study with chest X-Ray and lung CT scan images," *Pattern Recognit. Lett.* **129**, 271–278 (2020).
2. R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep Learning in Ultrasound Imaging," *Proc. IEEE* **108**(1), 11–29 (2020).
3. S. Vedula, O. Senouf, A. M. Bronstein, O. V. Michailovich, and M. Zibulevsky, "Towards CT-quality Ultrasound Imaging using Deep Learning," *arXiv*, ArXiv171006304 Phys. (2017).
4. Y. H. Yoon and J. C. Ye, "Deep Learning for Accelerated Ultrasound Imaging," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2018), pp. 6673–6676.
5. M. Grewal, M. M. Srivastava, P. Kumar, and S. Varadarajan, "RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (2018), pp. 281–284.
6. N. Yamato, H. Niioka, J. Miyake, and M. Hashimoto, "Improvement of nerve imaging speed with coherent anti-Stokes Raman scattering rigid endoscope using deep-learning noise reduction," *Sci. Rep.* **10**(1), 15212 (2020).
7. S. Wang, B. Lin, G. Lin, R. Lin, F. Huang, W. Liu, X. Wang, X. Liu, Y. Zhang, and F. Wang, "Automated label-free detection of injured neuron with deep learning by two-photon microscopy," *J. Biophotonics* **13**, e201960062 (2020).
8. N. Vogler, A. Medyukhina, I. Latka, S. Kemper, M. Böhm, B. Dietzek, and J. Popp, "Towards multimodal nonlinear optical tomography – experimental methodology," *Laser Phys. Lett.* **8**(8), 617–624 (2011).
9. W. Becker, "Fluorescence lifetime imaging – techniques and applications," *J. Microsc.* **247**(2), 119–136 (2012).
10. V. B. Pelegati, J. Adur, A. A. De Thomaz, D. B. Almeida, M. O. Baratti, L. A. L. A. Andrade, F. Botcher-luiz, and C. L. Cesar, "Harmonic optical microscopy and fluorescence lifetime imaging platform for multimodal imaging," *Microsc. Res. Tech.* **75**(10), 1383–1394 (2012).
11. C. A. Patil, N. Bosschaart, M. D. Keller, T. G. van Leeuwen, and A. Mahadevan-Jansen, "Combined Raman spectroscopy and optical coherence tomography device for tissue characterization," *Opt. Lett.* **33**(10), 1135–1137 (2008).
12. P. C. Ashok, B. B. Praveen, N. Bellini, A. Riches, K. Dholakia, and C. S. Herrington, "Multi-modal approach using Raman spectroscopy and optical coherence tomography for the discrimination of colonic adenocarcinoma from normal colon," *Biomed. Opt. Express* **4**(10), 2179–2186 (2013).
13. A. T. Yeh, B. Kao, W. G. Jung, Z. Chen, J. Stuart Nelson, and B. J. Tromberg, "Imaging wound healing using optical coherence tomography and multiphoton microscopy in an in vitro skin-equivalent tissue model," *J. Biomed. Opt.* **9**(2), 248–253 (2004).
14. N. Iftimia, R. D. Ferguson, M. Mujat, A. H. Patel, E. Z. Zhang, W. Fox, and M. Rajadhyaksha, "Combined reflectance confocal microscopy/optical coherence tomography imaging for skin burn assessment," *Biomed. Opt. Express* **4**(5), 680–695 (2013).
15. K. Kong, C. J. Rowlands, S. Varma, W. Perkins, I. H. Leach, A. A. Koloydenko, H. C. Williams, and I. Nottingher, "Diagnosis of tumors during tissue-conserving surgery with integrated autofluorescence and Raman scattering microscopy," *Proc. Natl. Acad. Sci.* **110**(38), 15189–15194 (2013).
16. N. Vogler, S. Heuke, T. W. Bocklitz, M. Schmitt, and J. Popp, "Multimodal Imaging Spectroscopy of Tissue," *Annu. Rev. Anal. Chem.* **8**(1), 359–387 (2015).
17. M. Beeres, J. L. Wichmann, J. Paul, E. Mbalisike, M. Elsabaie, T. J. Vogl, and N.-E. A. Nour-Eldin, "CT chest and gantry rotation time: does the rotation time influence image quality?" *Acta Radiol.* **56**(8), 950–954 (2015).
18. J. Huang, Y. Cao, J. Wang, A. Liu, Q. Wu, Z. Chang, Z. Li, Y. Luo, L. Gao, and G. Yin, "Time-stretch-based multidimensional line-scan microscopy," *Opt. Lasers Eng.* **160**, 107197 (2023).
19. N. Goel, A. Yadav, and B. M. Singh, "Medical image processing: A review," in *2016 Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with Their Impact on Humanity (CIPECH)* (2016), pp. 57–62.



20. S. V. M. Sagheer and S. N. George, "A review on medical image denoising algorithms," *Biomed. Signal Process. Control* **61**, 102036 (2020).
21. JNTUH University, Telangana, India and also Dept of ECE, S R Engineering College (Autonomous), Warangal, India, S. Kollem, K. R. L. Reddy, and D. S. Rao, "A Review of Image Denoising and Segmentation Methods Based on Medical Images," *Int. J. Mach. Learn. Comput.* **9**(3), 288–295 (2019).
22. V. Mannam, Y. Zhang, Y. Zhang, Y. Zhu, E. Nichols, Q. Wang, V. Sundaresan, S. Zhang, C. Smith, P. W. Bohn, P. W. Bohn, S. S. Howard, and S. S. Howard, "Real-time image denoising of mixed Poisson–Gaussian noise in fluorescence microscopy images using ImageJ," *Optica* **9**(4), 335–345 (2022).
23. R. W. Gerchberg and W. O. Saxton, "Comment on 'A method for the solution of the phase problem in electron microscopy'," *J. Phys. Appl. Phys.* **6**(5), 101L31 (1973).
24. G. Yang, B. Dong, B. Gu, J. Zhuang, and O. K. Ersoy, "Gerchberg–Saxton and Yang–Gu algorithms for phase retrieval in a nonunitary transform system: a comparison," *Appl. Opt.* **33**(2), 209–218 (1994).
25. J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.* **21**(15), 2758–2769 (1982).
26. F. Fogel, I. Waldspurger, and A. d'Aspremont, "Phase retrieval for imaging problems," *Math. Program. Comput.* **8**(3), 311–335 (2016).
27. G. Whyte and J. Courtial, "Experimental demonstration of holographic three-dimensional light shaping using a Gerchberg–Saxton algorithm," *New J. Phys.* **7**, 117 (2005).
28. K. Weiss, T. M. Khoshgofaar, and D. Wang, "A survey of transfer learning," *J. Big Data* **3**(1), 9 (2016).
29. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Trans. Image Process.* **26**(7), 3142–3155 (2017).
30. J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning Image Restoration without Clean Data," *arXiv arXiv.1803.04189* (2018).
31. S. W. Zamir, A. Arora, S. H. Khan, H. Munawar, F. S. Khan, M.-H. Yang, and L. Shao, "Learning Enriched Features for Fast Image Restoration and Enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(2), 1934–1948 (2023).
32. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning Enriched Features for Real Image Restoration and Enhancement," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds., Lecture Notes in Computer Science (Springer International Publishing, 2020), Vol. 12370, pp. 492–511.
33. C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *arXiv, ArXiv150100092 Cs* (2015).
34. C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016).
35. C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2015), pp. 1–9.
36. L. Tan and J. Jiang, "Chapter 13 - Image Processing Basics," in *Digital Signal Processing (Third Edition)*, L. Tan and J. Jiang, eds. (Academic Press, 2019), pp. 649–726.
37. Y. Gao and L. Cao, "A Complex Constrained Total Variation Image Denoising Algorithm with Application to Phase Retrieval," **11** (n.d.).
38. H. Chang, Y. Lou, Y. Duan, and S. Marchesini, "Total Variation–Based Phase Retrieval for Poisson Noise Removal," *SIAM J. Imaging Sci.* **11**(1), 24–55 (2018).
39. O. Oh, Y. Kim, D. Kim, D. S. Hussey, and S. W. Lee, "Phase retrieval based on deep learning in grating interferometer," *Sci. Rep.* **12**(1), 6739 (2022).
40. Ç. Işıl, F. S. Oktem, and A. Koç, "Deep iterative reconstruction for phase retrieval," *Appl. Opt.* **58**(20), 5422–5431 (2019).
41. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv, ArXiv14091556 Cs* (2015).
42. S. Heuke, N. Vogler, T. Meyer, D. Akimov, F. Kluschke, H. J. Rowert-Huber, J. Lademann, B. Dietzek, and J. Popp, "Detection and Discrimination of Non-Melanoma Skin Cancer by Multimodal Imaging," *Healthcare* **1**(1), 64–83 (2013).
43. J. Xiao, Z. Liu, P. Zhao, Y. Li, and J. Huo, "Deep Learning Image Reconstruction Simulation for Electromagnetic Tomography," *IEEE Sens. J.* **18**(8), 3290–3298 (2018).